Calcif Tissue Int (2005) 77:281–290
DOI: 10.1007/s00223-005-0039-2

Calcified Tissue International

Compartmental Bone Morphometry in the Mouse Femur: Reproducibility and Resolution Dependence of Microtomographic Measurements

T. Kohler, M. Beyeler, D. Webster, R. Müller

Institute for Biomedical Engineering, Swiss Federal Institute of Technology (ETH) and University of Zürich, Moussonstrasse 18, 8044 Zürich, Switzerland

Received: 15 July 2005 / Accepted: 22 July 2005 / Online publication: 4 November 2005

Abstract. Microcomputed tomography (μCT) is widely used for nondestructive bone phenotyping in small animals, especially in the mouse. Here, we investigated the reproducibility and resolution dependence of μCT analysis of microstructural parameters in three different compartments in the mouse femur. Reproducibility was assessed with respect to precision error ($PE_{\%CV}$) and intraclass correlation coefficient (ICC). We examined 14 left femurs isolated postmortem from two strains of mice (seven per group). Measurements and analyses were repeated five times on different days. In a second step, analysis was repeated again five times for a single measurement. Resolution dependence was assessed by high-resolution measurements (10 μm) in one strain and subsequent image degrading. Reproducibility was better in full bone compartment and in cortical bone compartment in the diaphysis ($PE_{\%CV}$ = 0.06–2.16%) than in trabecular compartment in the distal metaphysis ($PE_{\%CV}$ = 0.59–5.24%). Nevertheless, ICC (0.92–1.00) showed a very high reliability of the assessed parameters in all regions, indicating very small variances within repeated measurements compared to the population variances. Morphometric indices computed from lower- and higher-resolution images displayed in general only weak dependence and were highly correlated with each other (R^2 = 0.91–0.99). The results show that parameters in the full and cortical compartments were very reproducible, whereas precision in the trabecular compartment was somewhat lower. Nevertheless, all compartmental analysis methods were very robust, as shown by the high ICC values, demonstrating high suitability for application in inbred strains, where highest precision is needed due to small population variances.

Key words: Microcomputed tomography (μCT) — Inbred mouse strain — Bone density — Bone architecture/structure — Reproducibility — Resolution dependence

The use of the mouse as a model for human musculoskeletal diseases has increased in popularity as the mouse genome has been well characterized. The advantages of the mouse model are that various strains have been observed to exhibit disease state characteristics similar to those found in humans and that the mouse is easily accessible to manipulation of the genetic makeup by either gene knockout, gene overexpression (transgenes), or genetic breading strategies [1]. Despite the wide variety of methods available today for assessing bone properties, nondestructive imaging methods such as peripheral quantitative computed tomography (pQCT) and dual X-ray absorptiometry (DXA) have limited use for evaluation of microstructural parameters due to the small size of murine bones. Histomorphometry, despite its very high resolution, is a destructive and time-consuming method. Alternatively, microcomputed tomography (μCT) is fully nondestructive and well suited for assessing truly three-dimensional (3D) microstructural bone properties [2–7]. Previous studies have shown μCT to be an accurate technique with close correlations between microtomographic and histomorphometric measurements of static structural bone metrics in various applications [8–12]. However, accuracy of μCT is always dependent on the resolution chosen compared to the structure being measured and has therefore to be investigated in all the different resolution/application configurations.

Reproducibility has been examined for different analysis procedures and tissue parameters for various bone measurement techniques including DXA [13–15], ultrasound (US) [16, 17], magnetic resonance imaging (MRI) [18–21], and pQCT [12, 22]. However, reproducibility data for structural bone properties analyzed by means of μCT are quite rare and, to our knowledge, have not been assessed so far in the mouse.

In this study, we investigated the reproducibility of microtomographic analysis of structural parameters in three different compartments of the mouse femur and for two different strains. As measures of reproducibility, we examined the precision error (PE) as introduced by Gluer et al. [23] and the intraclass correlation coefficient (ICC) as first described by Shrout [24]. We also tested

Correspondence to: R. Müller; E-mail: ralph.mueller@ethz.ch

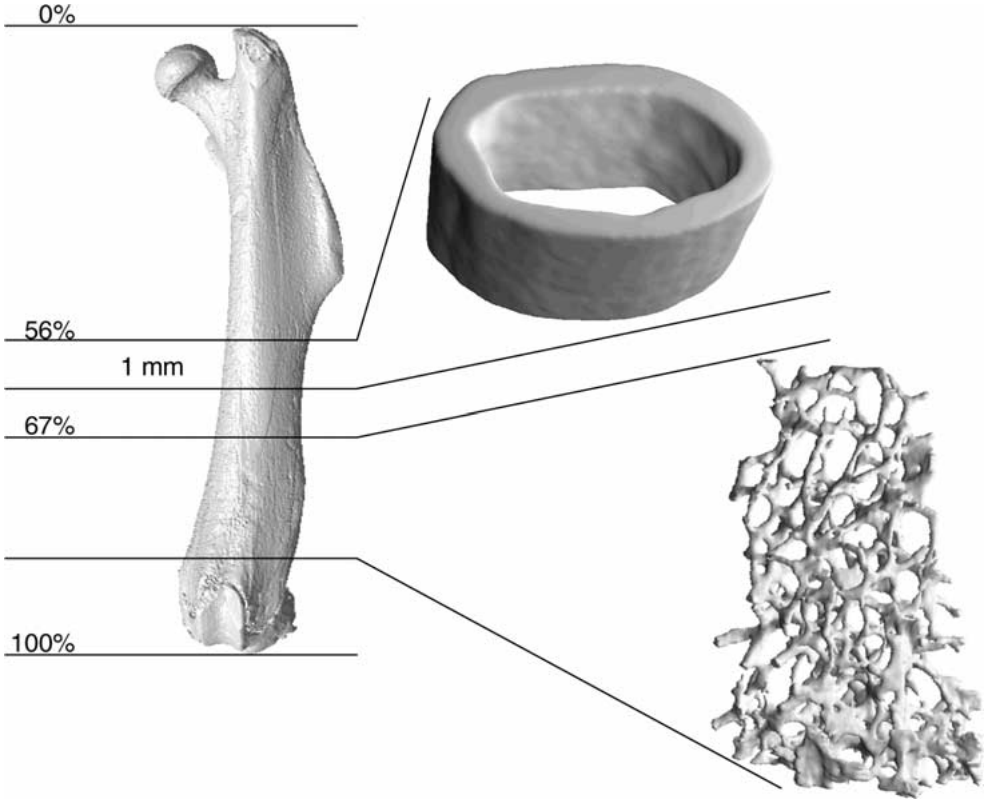

Fig. 1. Compartmental mouse femur analysis. Full femur compartment is shown on the left side, cortical and trabecular compartment on the right side. Percentages indicate the position of the compartments.

whether the chosen scanning configuration provides sufficient accuracy to represent the microarchitecture of murine trabecular bone. Therefore, we investigated the resolution dependence of microstructural parameters in a subvolume of the trabecular bone compartment.

Materials and Methods

Animals

For the study, mice of two different inbred strains were obtained from Harlan (Jerusalem, Israel). One strain was C3H, which is typically referred to as a high-bone mass strain [25, 26], and the other strain was SJL, which exhibits thin bone cortices and low trabecular volume density. Two groups including seven female animals, 8–11 weeks old, were formed for each strain. At the time of death, left intact femurs were excised from all animals and kept in 70% ethanol for further analysis. Experimental protocols were approved by the Institutional Animal Care and Use Committee, Faculty of Medicine, Hebrew University of Jerusalem, Israel.

Experimental Protocol

For testing the reproducibility of the 3D morphometric evaluation (measurement and analysis), the left femur of each mouse in the two groups was measured five times on different days on a μCT imaging system with 20 μm isotropic voxel resolution and analyzed by a single operator. In a second setup, the reproducibility of the compartmental analysis procedure only was investigated. Therefore, the median measurement according to trabecular bone volume density was taken and reanalyzed five times, again by a single operator. Additionally, we measured all seven femurs of the SJL strain with 10 μm isotropic voxel resolution and digitally reduced voxel resolution to 20, 30, and 40 μm voxel size to investigate the resolution dependence of the structural parameters in trabecular bone.

Microtomographic Imaging System

The microtomographic imaging system (μCT 40; Scanco Medical, Bassersdorf, Switzerland) used in this study was equipped with a 5 μm focal spot X-ray tube as a source. A 2D charge-coupled device, coupled to a thin scintillator as a detector, permitted acquisition of 20 tomographic images in parallel. The long axis of the femur was orientated orthogonal to the axis of the X-ray beam. The X-ray tube was operated at 50 kVp and 160 μA. The integration time was set to 100 milliseconds. Scans for testing reproducibility were performed at a nominal resolution of 20 μm in all three spatial dimensions (medium resolution mode). Two-dimensional CT images were reconstructed in 1024×1024 pixel matrices from 1,000 projections using a standard convolution-backprojection procedure with a Shepp and Logan filter. Images were stored in 3D arrays with an isotropic voxel size of 20 μm. Scans for testing resolution dependence were performed with the same positioning of the femurs as in the reproducibility scans but at a nominal resolution of 10 μm (high-resolution mode). The same standard convolution-backprojection procedure was used, and images were reconstructed in 2048×2048 pixel matrices.

Compartment Definitions: Volumes of Interest

The measurements for testing reproducibility were then analyzed in three different preselected analysis compartments [11]: compartement I included the full femur, compartment II contained the trabecular bone in the distal metaphysis, and

compartment III comprised a 1-mm-thick slab in the midshaft (Fig. 1). These volumes of interest (VOI) had to be defined for each measurement and analysis. To generate the compartments, two contours had to be drawn slice-wise by an operator in 2D using a semiautomatic tracking algorithm [27]. Contours were drawn on slices orthogonal to the long axis of the femur (superior-inferior direction). The first contour characterized the outer envelope of the femur, which was then used to define the full femur (compartment I). Additionally, this outer contour was used to automatically create the cortical compartment III, which started at 56% of the whole femur length (calculated from the great trochanter) and contained 50 slices resulting in a stack height of 1 mm. The second contour described the inner contour delineating cortical from trabecular bone. This contour defining the trabecular bone compartment in the distal metaphysis started at two-thirds of the femur length and stopped at a defined number of slices above the growth plate, to be sure to exclude any cortical structures from the growth plate. The number of slices that have to be subtracted from the first slice showing parts of the growth plate is directly proportional to the length of the femur and represents 30 slices (0.6 mm) for a normal adult mouse femur of 800 slices (16 mm). Once the two contours were drawn, all three compartments were generated fully automatically using DCL (Hewlett-Packard, Palo Alto, CA) and IPL (Scanco Medical) script files.

Resolution dependence of microstructural parameters was investigated in the trabecular bone compartment in the distal metaphysis. Within that volume we selected a cube with a side length of 840 μm (84 voxel) as the VOI. For the digitally reduced images (20, 30, 40 μm), we also resampled this VOI to voxel side lengths of 63, 42, and 21, respectively, though keeping the volume and the position of the VOI constant.

Image Processing

A constrained 3D Gaussian filter was used to partly suppress the noise in the volumes, and the mineralized tissue was segmented from soft tissues by a global thresholding procedure [2]. These segmentation steps were applied to all three analysis compartments but with different parameters. As values for the Gaussian filter, we used a sigma of 0.8 and support of one voxel for the full and trabecular femur and a sigma of 3 and support of three voxels for the cortical bone compartment (elimination of small cortical porosities). The threshold values were then set to 22.4% of the maximum gray scale value for the full femur analysis, 16.0% for the trabecular compartment, and 25.0% for the cortical bone compartment. For resolution dependence, we first segmented the 10 μm resolution images as described above and applied the resolution reduction on the binarized images. Resolution reduction was performed with factors of 2, 3, and 4, resulting in images with isotropic voxel sizes of 20, 30, and 40 μm, respectively.

3D Morphometric Parameter Evaluation

Morphometric parameters were determined using a direct 3D approach [28] in each of the three different analysis compartments. For the whole bone, only apparent volume density (AVD) was assessed, which is the number of bone voxels, defined by the thresholding procedure, divided by the number of all voxels within the outer contour describing the bone envelope. Parameters determined in the metaphyseal trabecular bone included bone volume density (BV/TV), trabecular thickness (Tb.Th), trabecular spacing (Tb.Sp), trabecular number (Tb.N), and connectivity density (Conn.D). To assess resolution dependence, the same structural parameters (BV/TV, Tb.Th, Tb.Sp, Tb.N, and Conn.D) were also determined in the trabecular subvolumes of the different resampled images. Percent cortical volume (%BV) and cortical thickness (Ct.Th) were assessed in the 1-mm-thick cortical volume in the diaphysis.

Statistics

PEs were defined to best characterize the reproducibility of a given bone measurement technique [23]. In that approach, the short-term precision of an individual subject (SD_j) is defined as the standard deviation of n repeated measurements on a given subject j. PEs may be expressed in absolute numbers or as coefficients of variation ($\%CV_j$) of repeated measurements given on a percentage basis. To determine the short-term precision of the technique under investigation, the individual precision data have to be pooled. Since only the measured individual variances (SD_j^2) can be considered an unbiased estimate of the parameter σ^2 of the Gaussian normal distribution, these individual variances have to be averaged arithmetically to get the technique's squared precision errors, SD^2 and $\%CV^2$, respectively. Consequently, the technique's PE is best given by the root mean-square (RMS) average of the individual subject's PEs for each of the m subjects:

$$SD = \sqrt{\sum_{j=1}^{m}\sum_{i=1}^{n_j}\frac{\left(x_{ij}-\overline{x_j}\right)^2}{df}} \qquad \text{(Eq. 1)}$$

$$\%CV = \sqrt{\sum_{j=1}^{m}\sum_{i=1}^{n_j}\frac{\left(x_{ij}-\overline{x_j}\right)^2}{df\cdot\overline{x_j}^2}}\cdot 100\% \qquad \text{(Eq. 2)}$$

The result of the ith measurement for subject j is x_{ij}, and $\overline{xj}$ is the mean of all x_{ij} for subject j. The total number of degrees of freedom (df) for a technique is the sum of the degrees of freedom (df_j) of the measurements in the individual subjects:

$$df = \sum_{j=1}^{m} df_j = \sum_{j=1}^{m}(n_j - 1) \qquad \text{(Eq. 3)}$$

In this study, we calculated the SD PEs (PE_{SD}) as well as the %CV PE ($PE_{\%CV}$) for the above-described eight structural parameters for the scheme where we investigated repeated measurements and analyses as well as repeated analyses on a single measurement, respectively.

To determine how accurate the PEs were, we calculated a confidence interval (CI) for each of the $PE_{\%CV}$ values. The correct estimate of the 95% CI is not given by ± 2 times the observed SD of all of the individual subjects' PEs. Since a PE cannot become negative but has no upper limit, the CI has to be asymmetric. Therefore, the chi-squared (χ^2) distribution was used. This distribution is asymmetric and depends on the total number of degrees of freedom (df) given by Eq. 3. The formula for a $(1-\alpha)\cdot 100\%$ CI of the true precision error σ is as follows:

$$\frac{df}{\chi^2_{1-\frac{\alpha}{2},df}}\%CV^2 < \sigma^2 < \frac{df}{\chi^2_{\frac{\alpha}{2},df}}\%CV^2 \qquad \text{(Eq. 4)}$$

Additionally, the mean of each parameter for each mouse strain used to determine the precision was calculated. Those mean values should be stated in addition to the PEs because the precision may not be the same in osteopenic or osteoporotic populations as it is in normal populations [29].

Another term classically used in reproducibility studies is ICC, described by Shrout [24] and discussed more recently by Vargha [30]. ICC is defined as the intersubject variance divided by the population variance. It therefore varies between 0 and 1, where 1 indicates perfect reproducibility. The ICC model, chosen for a reproducibility study with given repeated measurements, is the two-way mixed model. This model assumes randomly selected subjects, which are measured with a fixed number of repetitions, where those repetitions are the only ones of interest. Following Shrout [24], the ICC(3,1), which corresponds to the two-way mixed model, may be calculated as follows:

Table 1. PEs and CIs for repeated measurement and analysis

Parameters	SJL				C3H			
	Mean	PE_{SD}	$PE_{\%CV}$	95% CI $PE_{\%CV}$	Mean	PE_{SD}	$PE_{\%CV}$	95% CI $PE_{\%CV}$
AVD (%)	62.442	0.309	0.50%	0.40–0.68%	65.982[a]	0.536	0.81%	0.64–1.11%
BV/TV (%)	21.367	0.554	2.75%	2.17–3.74%	38.932[a]	1.406	3.56%	2.81–4.84%
Tb.Th (mm)	0.093	0.002	1.81%	1.43–2.46%	0.099	0.003	2.79%	2.21–3.80%
Tb.Sp (mm)	0.269	0.005	1.63%	1.29–2.21%	0.163[a]	0.003	1.88%	1.49–2.56%
Tb.N (mm^{-1})	3.754	0.063	1.78%	1.41–2.42%	5.758[a]	0.091	1.58%	1.25–2.15%
Conn.D (mm^{-3})	57.041	2.835	5.24%	4.14–7.12%	117.611[a]	4.472	3.80%	3.00–5.17%
%BV (%)	41.647	0.791	1.86%	1.47–2.53%	59.498[a]	0.742	1.28%	1.01–1.74%
Ct.Th (mm)	0.169	0.004	2.16%	1.70–2.93%	0.250[a]	0.003	1.37%	1.08–1.87%

[a] Student's *t*-test between SJL and C3H, $P < 0.05$

$$ICC = \frac{F_0 - 1}{F_0 + (n-1)} \quad \text{(Eq. 5)}$$

where F_0 is the ratio of between-subject mean squares and the residual within-subject mean squares and n is the number of repetitions. We calculated the ICC and its 95% CI again for all evaluated parameters in both repetition schemes. The formula for calculating the CIs of the ICC can be found in Shrout [24].

To investigate the difference in the parameters of interest between the two groups (mouse strains) and the possible effects within the repeated measurements and analysis, we performed the general linear model (GLM) repeated measures procedure, which provides an analysis of variance when the same measurement is made several times on each subject.

Statistics were analyzed with the SPSS software package (version 12.0 for Windows; SPSS, Chicago, IL).

Results

Repeated Measurement and Analysis

The GLM repeated measures procedure showed a significant effect between the two strains for all calculated parameters except Tb.Th. Therefore, the PEs had to be calculated for both groups because, as mentioned before and described by Bonnick [29], they might be different for the two strains with different bone characteristics.

Then the 95% CI for this experimental setup was calculated according to Eq. 3 and Eq. 4. The 95% CI ranges from −21% to +36% of the PEs. The $PE_{\%CV}$ for the SJL strain was in the range from 0.5% (AVD) to 5.2% (Conn.D) and for the C3H strain, from 0.8% (AVD) to 3.8% (Conn.D). Means, PE_{SD}, and $PE_{\%CV}$ as well as the 95% CIs for $PE_{\%CV}$ are given in Table 1 and $PE_{\%CV}$ is graphically illustrated in Figure 2.

The ICC for the SJL strain ranged from 0.971 (Tb.Th) to 0.998 (Tb.Sp) and for the C3H strain, from 0.923 (Conn.D) to 0.996 (Ct.Th). ICC values and their corresponding 95% CIs can be found in Table 2.

The GLM repeated measures procedure showed in the multivariate analysis a significant ($P < 0.05$) within-subject effect. When we examined the univariate tests for each parameter, only AVD showed a significant effect (Fig. 3).

Repeated Analysis

When one selected measurement was reanalyzed again five times, the significant difference between the strains, as expected, remained the same for all parameters except Tb.Th, which already was not affected by strain. Values of $PE_{\%CV}$ ranged from 0.06% (%BV) to 3.48% (Tb.Th) in the SJL strain and from 0.08% (%BV) to 2.83% (BV/TV) in the C3H strain. Means, PE_{SD}, and $PE_{\%CV}$ as well as the 95% CI for $PE_{\%CV}$ are given in Table 3. A graphical illustration of the $PE_{\%CV}$ can be found in Figure 2.

The ICC for the SJL strain ranged from 0.920 (Tb.Th) to 1.000 (AVD, Tb.Sp, %BV, Ct.Th) and for the C3H strain, from 0.976 (Conn.D) to 1.000 (%BV, Ct.Th). ICC values and their corresponding 95% CIs for the repeated analysis setup can be found in Table 4.

Like in the repeated measurement and analysis setup, we found also for the repeated analysis a significant effect between the repetitions. Looking at the univariate tests revealed as before no significant effect in AVD but a mild effect in both BV/TV and Tb.Th ($P < 0.05$).

Resolution Dependence

From the analysis of resolution dependence, we found resolution having a strong effect on Conn.D. The difference in Conn.D between 10 and 20 μm resolution was 14.5%. Changes for BV/TV, Tb.Th, Tb.Sp, and Tb.N were small, ranging from 0.05% for BV/TV to 3.1% for Tb.N. Nevertheless, all structural parameters were highly correlated between the 10 and 20 μm resolution images: BV/TV ($R^2 = 0.99$), Tb.Th ($R^2 = 0.91$), Tb.Sp ($R^2 = 0.99$), Tb.N ($R^2 = 0.99$), and Conn.D ($R^2 = 0.98$). Values for structural parameters in all analyzed resolution modes can be found in Table 5. A graphical representation of the influence of resolution on trabecular bone structure is shown in Figure 4.

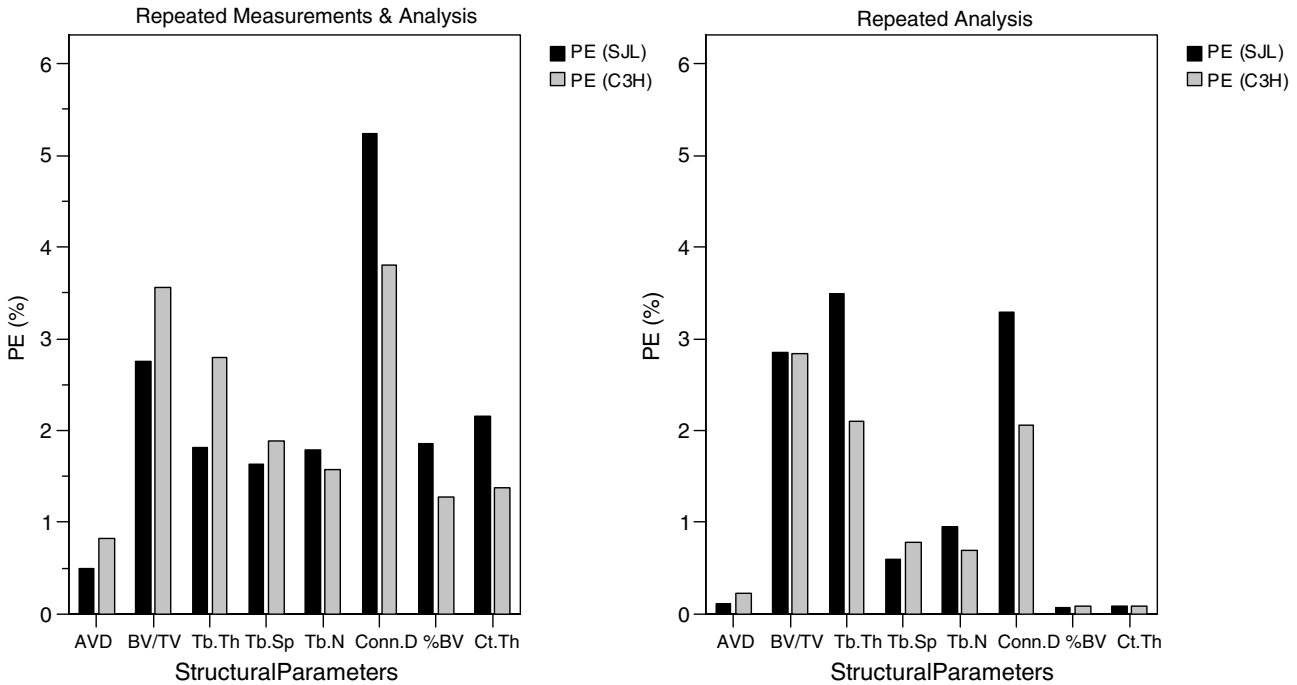

Fig. 2. $PE_{\%CV}$ for the repeated measurements and analysis on the left side and the repeated analysis only on the right side for two strains of mice and all structural parameters.

Table 2. ICCs and CIs for repeated measurement and analysis

	SJL		C3H	
Parameters	ICC	95% CI ICC	ICC	95% CI ICC
AVD (%)	0.997	0.990–0.999	0.994	0.983–0.999
BV/TV (%)	0.997	0.992–0.999	0.988	0.964–0.998
Tb.Th (mm)	0.971	0.914–0.994	0.976	0.928–0.995
Tb.Sp (mm)	0.998	0.993–1.000	0.994	0.982–0.999
Tb.N (mm^{-1})	0.997	0.991–0.999	0.990	0.969–0.998
Conn.D (mm^{-3})	0.989	0.967–0.998	0.923	0.770–0.985
%BV (%)	0.972	0.915–0.994	0.994	0.981–0.999
Ct.Th (mm)	0.982	0.947–0.997	0.996	0.990–0.999

Discussion

In this study, we evaluated the reproducibility of μCT measurements in femurs of two strains of mice. We found that μCT has a high precision in evaluating bone parameters in mouse femurs. The PEs were larger for evaluated parameters in the metaphysis than for parameters describing the whole bone or the midshaft. The whole bone region was defined by the outer contour, which was also used to define the boundary of the cortical compartment. The built-in semiautomatic contouring algorithm was used to define the bone envelope, allowing minimal operator interaction and decision. Our data analysis showed that the definition of those two compartments was very reproducible. Values of $PE_{\%CV}$ in the repeated analysis were very low and ranged from 0.06% (%BV in SJL) to 0.23% (AVD in C3H). The PEs in these two compartments were considerably higher for the repeated measurement and analysis setup (Fig. 2). This can be explained by the repositioning of the femurs in the scanner causing angular variations of the selected compartments and a variation in both the defined contours and the gray scale distribution in the images. Nevertheless, with typical values for reproducibility around and below the 2% level, structural parameters in these compartments were still highly reproducible.

The trabecular compartment is defined by the inner contour. Since drawing this inner contour requires more operator interaction, it is not surprising that the PEs in this compartment are higher than in the regions defined by the outer contour. Additionally, selection of the trabecular compartment is, similar to the other re-

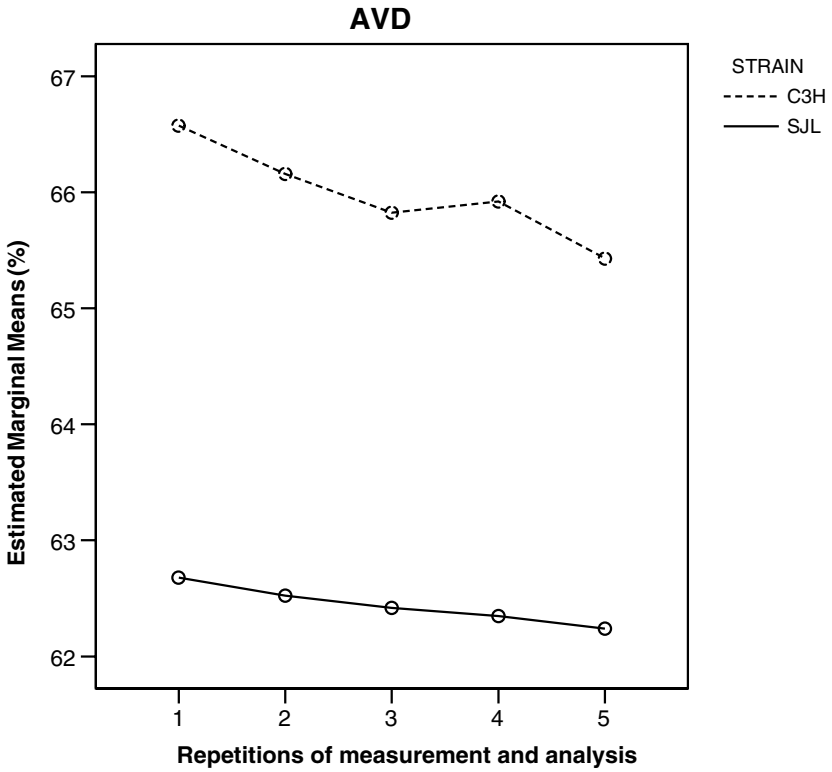

Fig. 3. Estimated marginal means calculated by GLM repeated measures analysis. Data for both strains show a significant trend ($P < 0.05$) for the five subsequent repetitions.

Table 3. PEs and CIs for repeated analysis

	SJL				C3H			
Parameters	Mean	PE_{SD}	$PE_{\%CV}$	95% CI $PE_{\%CV}$	Mean	PE_{SD}	$PE_{\%CV}$	95% CI $PE_{\%CV}$
AVD (%)	62.724	0.070	0.11%	0.09–0.16%	*66.132	0.144	0.23%	0.18–0.31%
BV/TV (%)	22.294	0.662	2.84%	2.25–3.87%	*38.316	1.090	2.83%	2.24–3.86%
Tb.Th (mm)	0.096	0.003	3.48%	2.75–4.74%	0.098	0.002	2.09%	1.66–2.85%
Tb.Sp (mm)	0.267	0.001	0.59%	0.47–0.81%	*0.165	0.001	0.78%	0.61–1.06%
Tb.N (mm^{-1})	3.780	0.038	0.95%	0.75–1.29%	*5.735	0.040	0.69%	0.55–0.94%
Conn.D (mm^{-3})	58.339	1.879	3.29%	2.60–4.48%	*117.567	2.388	2.06%	1.63–2.80%
%BV (%)	41.278	0.026	0.06%	0.05–0.09%	*59.200	0.047	0.08%	0.06–0.11%
Ct.Th (mm)	0.169	0.000	0.08%	0.07–0.11%	*0.248	0.000	0.08%	0.07–0.11%

*Student's *t*-test between SJL and C3H, $P < 0.05$

gions, also dependent on the alignment of the femurs in the scanner. This was also corroborated by the fact that PEs were generally lower for the repeated analysis setup. Nevertheless, analysis methods seemed very robust and precise when applied by an experienced operator. Further improvement of the reproducibility in the trabecular compartment as well as in the full and cortical compartments could be achieved using fully automated VOI definition methods, where no user interaction would be required and which would compensate for the angular misalignment of the sample in the scanner.

The ICC showed a very high reproducibility, ranging 0.920–1.000. Since between-subject variances of inbred strain groups are by definition expected to be very small, the ICC values found in this study are very promising. The within-subject variances (measurement and analysis errors) are thus very small, and the presented compartmental analysis procedure is a highly reliable method for assessing structural bone parameters in the mouse femur. Investigating the different compartments, the ICC values were generally also lower in the trabecular compartment than in the full and cortical VOI and, hence, confirmed the findings for the PEs.

Table 4. ICCs and CIs for repeated analysis

	SJL		C3H	
Parameters	ICC	95% CI ICC	ICC	95% CI ICC
AVD (%)	1.000	0.999–1.000	0.999	0.998–1.000
BV/TV (%)	0.996	0.988–0.999	0.996	0.988–0.999
Tb.Th (mm)	0.920	0.760–0.984	0.992	0.975–0.998
Tb.Sp (mm)	1.000	0.999–1.000	0.999	0.998–1.000
Tb.N (mm^{-1})	0.999	0.997–1.000	0.998	0.994–1.000
Conn.D (mm^{-3})	0.995	0.984–0.999	0.976	0.929–0.995
%BV (%)	1.000	1.000–1.000	1.000	1.000–1.000
Ct.Th (mm)	1.000	1.000–1.000	1.000	1.000–1.000

Table 5. Resolution dependence of structural parameters in trabecular bone

	BV/TV (%)	R^2	Tb.Th (mm)	R^2	Tb.Sp (mm)	R^2	Tb.N (mm^{-1})	R^2	Conn.D (mm^{-3})	R^2
10 μm	21.089		0.071		0.227		4.279		96.779	
20 μm	21.099	0.99	0.072	0.91	0.225	0.99	4.410	0.99	82.723	0.98
30 μm	21.063	0.99	0.075	0.96	0.229	0.99	4.424	0.98	72.584	0.91
40 μm	21.101	0.98	0.079	0.34	0.233	0.95	4.328	0.98	56.225	0.64

R^2 values, correlation to 10 μm results (taken as the gold standard)

The significant effect between repetitions that occurred for AVD in the repeated measurement and analysis setup and for BV/TV and Tb.Th in the repeated analysis seemed to be less problematic since those within-repetition variances were much smaller compared to the group variances. This could also be shown by the very high ICC, which makes it possible to almost perfectly discriminate parameters between different animals. It is still interesting to know that due to the operator learning effect with drawing contours, time-dependent inaccuracies may arise. Therefore, it is absolutely mandatory to analyze the data of different groups in a random fashion and for a new operator to have an appropriate learning phase using test data before analyzing real studies. This is especially important for upcoming *in vivo* systems where significant although subtle time-dependent trends could be produced that would be caused solely by the operator learning curve.

Although the two groups showed a strongly significant difference in all analyzed parameters except Tb.Th, PE_{SD} and ICC were in the same range for both strains. Therefore, we expect that the precision, including the upper confidence limit for PEs and the lower confidence limit for ICCs, may be generalized to femurs of other mouse strains with similar bone characteristics (age, size).

PEs and ICCs can only be calculated for the analyzed parameters, which makes it difficult to compare them to other techniques since different techniques usually produce different parameters. Only few data are available using some of the same parameters as in this study. Schmidt et al. [12] assessed reproducibility in mouse femurs based on pQCT measurements. They found a $PE_{\%CV}$ of 2.1% for Ct.Th as measured from pQCT. This is in good accordance with the 1.3% and 2.2% PEs for C3H and SJL, respectively, found in our study using μCT. Both techniques and analysis methods seem to have a similar reproducibility for the Ct.Th parameter in mouse femurs. Müller et al. [22] investigated human trabecular structural indices (BV/TV, Tb.N, Tb.Th, and Tb.Sp) *in vivo* in the distal end of the radius using pQCT. In their study, ICC ranged from 0.453 (Tb.Th) to 0.995 (Tb.Sp). Except for the Tb.Th parameter, they also showed a similar reproducibility for pQCT-derived trabecular parameters in human cancellous bone as found in our study. Nevertheless, since only six subjects with two repetitions were examined, it has to be noted that the CIs were quite large.

Gomberg et al. [20] investigated reproducibility of human trabecular bone structural parameters assessed by high-resolution MRI (μMRI) and found a $PE_{\%CV}$ for BV/TV of 4.6% and an ICC of 0.97 in the radius and a $PE_{\%CV}$ of 6.1% and an ICC of 0.68 in the tibia. Newitt et al. [18] and Laib et al. [19] also investigated the reproducibility of human trabecular bone structure in the distal radius by means of μMRI. The 2D apparent trabecular parameters described by Newitt et al. [18] were slightly less reproducible than the true 3D parameters assessed by Laib et al. [19] but consistent overall with data shown by Gomberg et al. [20]. The trabecular compartment analysis presented here for μCT ($PE_{\%CV} < 3.6\%$ and ICC > 0.988 for BV/TV) seems to be even more reproducible with lower errors and higher ICCs.

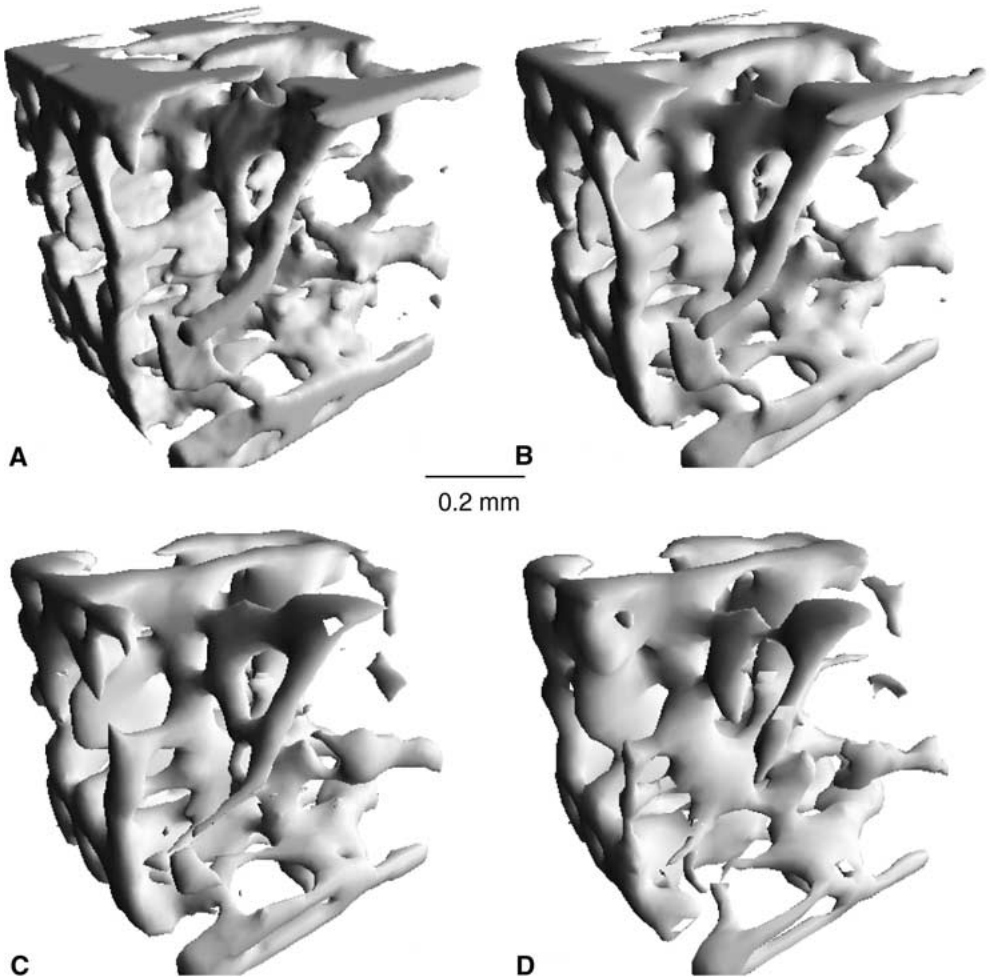

Fig. 4. Effect of resolution dependence on trabecular bone structure. **A** Original image measured with voxel resolution of 10 μm. **B-D** Trabecular structure with digitally reduced voxel resolution to 20, 30 and 40 μm,respectively.

With respect to resolution dependence, it must be noted that digital resampling of a high-resolution image at lower resolutions compared to actually scanning at low resolutions underestimates the resolution dependence of structural parameters [31]. Nevertheless, only digital resampling allows comparison of exactly the same VOI, whereas scanning at different resolutions brings in the issue of reproducibility of the analyzed volume because the VOI has to be selected again. As shown here, the PE might be up to 5% for structural parameters even in measurements with the same resolution. However, being aware of this effect, we chose to reduce the voxel size of already binarized images, which is a worst-case scenario because structures below the resampled resolution are preferentially lost. Voxel reduction on gray scale images tends to preserve structural integrity by making structural components thicker. The effect of losing very thin structures with lower resolution is best seen in the very strong decrease of Conn.D with resolution. Since Conn.D is very sensitive to resolution and showed the highest PE when measured with 20 μm resolution, it might be a problematic parameter; and we recommend being careful with its use at this resolution. Nevertheless, we conclude that the relatively weak sensitivity of all other parameters to the actual image resolution in the investigated range gives good confidence that the chosen imaging mode is sufficient to assess and describe the trabecular microstructure as a whole (see Fig. 4).

However, one of the problems with limited resolution imaging is that actual thickness measurements on the level of individual trabeculae will always be overestimated. This is partly due to the fact that the operator typically will choose a relatively low segmentation threshold to best preserve the connectivity and with that increase thickness and volume. Nevertheless, as shown in the resolution dependence study, the results from the different resolution steps (except 40 μm) are highly correlated with each other, indicating that group differences can still be preserved. Additionally, a very recent study [32] showed that 3D measurements of thickness are always about 40% higher than 2D results from classical histology. Finally, a third source of variation is due to the analysis of different anatomic regions

within the bone. This is actually best illustrated by the current study, where mean thickness for the integral metaphyseal area was 96 μm for SJL mice but thickness measures in the same mice from the central part of the metaphysis dropped to 72 μm.

It should be mentioned that we chose this imaging mode with nominal resolution of 20 μm because of the very fast scanning time (< 30 minutes) as well as the reasonable size of the generated image data (< 500 MB). These criteria allow for high-throughput scanning, which is a precondition for phenotypic characterization of large mouse populations (> 1,000 animals) in genetic linkage studies. Nevertheless, if very accurate results are needed, the highest achievable resolution should be used.

In conclusion, microtomographic imaging is a very reliable and precise method, well suited for the nondestructive analysis of microstructural bone properties, allowing assessment of even subtle skeletal phenotypes in different mouse strains as they often appear in mouse genetics and molecular intervention.

Acknowledgment. This study was supported by the SNF Professorship in Bioengineering of the Swiss National Science Foundation (FP 620-58097.99). The authors thank Dr. Itai Bab for providing the mice and Christoph Buser from the Seminar for Statistics (ETH Zürich) for the helpful discussions on statistics.

References

1. Silver LM (1995) Mouse genetics. Oxford University Press, New York
2. Rüegsegger P, Koller B, Müller R (1996) A microtomographic system for the nondestructive evaluation of bone architecture. Calcif Tissue Int 58:24–29
3. Müller R, Rüegsegger P (1997) Microtomographic imaging for the nondestructive evaluation of trabecular bone architecture. Stud Health Technol Inform 40:61–79
4. Eckstein F, Lochmüller EM, Koller B, Wehr U, Weusten A, Rambeck W, Hoeflich A, Wolf E (2002) Body composition, bone mass and microstructural analysis in GH-transgenic mice reveals that skeletal changes are specific to bone compartment and gender. Growth Horm IGF Res 12:116–125
5. Müller R (2003) Bone microarchitecture assessment: current and future trends. Osteoporos Int 14 (suppl 5):89–99
6. Borah B, Dufresne TE, Chmielewski PA, Johnson TD, Chines A, Manhart MD (2004) Risedronate preserves bone architecture in postmenopausal women with osteoporosis as measured by three-dimensional microcomputed tomography. Bone 34:736–746
7. Ito M, Nishida A, Nakamura T, Uetani M, Hayashi K (2002) Differences of three-dimensional trabecular microstructure in osteopenic rat models caused by ovariectomy and neurectomy. Bone 30:594–598
8. Müller R, Van Campenhout H, Van Damme B, Van Der Perre G, Dequeker J, Hildebrand T, Rüegsegger P (1998) Morphometric analysis of human bone biopsies: a quantitative structural comparison of histological sections and microcomputed tomography. Bone 23:59–66
9. Kapadia RD, Stroup GB, Badger AM, Koller B, Levin JM, Coatney RW, Dodds RA, Liang X, Lark MW, Gowen M (1998) Applications of microCT and MR microscopy to study pre-clinical models of osteoporosis and osteoarthritis. Technol Health Care 6:361–372
10. Balto K, Muller R, Carrington DC, Dobeck J, Stashenko P (2000) Quantification of periapical bone destruction in mice by microcomputed tomography. J Dent Res 79:35–40
11. Alexander JM, Bab I, Fish S, Müller R, Uchiyama T, Gronowicz G, Nahounou M, Zhao Q, White DW, Chorev M, Gazit D, Rosenblatt M (2001) Human parathyroid hormone 1-34 reverses bone loss in ovariectomized mice. J Bone Miner Res 16:1665–1673
12. Schmidt C, Priemel M, Kohler T, Weusten A, Müller R, Amling M, Eckstein F (2003) Precision and accuracy of peripheral quantitative computed tomography (pQCT) in the mouse skeleton compared with histology and microcomputed tomography (microCT). J Bone Miner Res 18:1486–1496
13. Lochmüller EM, Jung V, Weusten A, Wehr U, Wolf E, Eckstein F (2001) Precision of high-resolution dual energy X-ray absorptiometry of bone mineral status and body composition in small animal models. Eur Cell Mater 1:43–51
14. Henzell S, Dhaliwal S, Pontifex R, Gill F, Price R, Retallack R, Prince R (2000) Precision error of fan-beam dual X-ray absorptiometry scans at the spine, hip, and forearm. J Clin Densitom 3:359–364
15. Iida-Klein A, Lu SS, Yokoyama K, Dempster DW, Nieves JW, Lindsay R (2003) Precision, accuracy, and reproducibility of dual X-ray absorptiometry measurements in mice in vivo. J Clin Densitom 6:25–33
16. Leong KH (1997) Ultrasound of the tibia – precision error, left versus right sides and correlation with bone mineral density. Ann Acad Med Singapore 26:747–749
17. Giraudeau B, Gomez MA, Defontaine M (2003) Assessing the reproducibility of quantitative ultrasound parameters with standardized coefficient of variation or intraclass correlation coefficient: a unique approach. Osteoporos Int 14:614–615
18. Newitt DC, van Rietbergen B, Majumdar S (2002) Processing and analysis of in vivo high-resolution MR images of trabecular bone for longitudinal studies: reproducibility of structural measures and microfinite element analysis derived mechanical properties. Osteoporos Int 13:278–287
19. Laib A, Newitt DC, Lu Y, Majumdar S (2002) New model-independent measures of trabecular bone structure applied to in vivo high-resolution MR images. Osteoporos Int 13:130–136
20. Gomberg BR, Wehrli FW, Vasilic B, Weening RH, Saha PK, Song HK, Wright AC (2004) Reproducibility and error sources of microMRI-based trabecular bone structural parameters of the distal radius and tibia. Bone 35:266–276
21. Kubo H, Harada M, Sakama M, Nishitani H (2003) Reproducibility of metabolite concentration evaluated by intraclass correlation coefficient using clinical MR apparatus. J Comput Assist Tomogr 27:449–453
22. Müller R, Hildebrand T, Hauselmann HJ, Rüegsegger P (1996) In vivo reproducibility of three-dimensional structural properties of noninvasive bone biopsies using 3D-pQCT. J Bone Miner Res 11:1745–1750
23. Gluer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK (1995) Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. Osteoporos Int 5:262–270
24. Shrout PE (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428
25. Judex S, Garman R, Squire M, Donahue LR, Rubin C (2004) Genetically based influences on the site-specific regulation of trabecular and cortical bone morphology. J Bone Miner Res 19:600–606
26. Beamer WG, Donahue LR, Rosen CJ, Baylink DJ (1996) Genetic variability in adult bone density among inbred strains of mice. Bone 18:397–403
27. Rüegsegger P, Münch B, Felder M (1993) Early detection of osteoarthritis by 3D computed tomography. Technol Health Care 1:53–66

28. Hildebrand T, Laib A, Müller R, Dequeker J, Rüegsegger P (1999) Direct three-dimensional morphometric analysis of human cancellous bone: microstructural data from spine, femur, iliac crest, and calcaneus. J Bone Miner Res 14:1167–1174
29. Bonnick SL (2004) Bone densitometry in clinical practice, 2nd ed. Humana Press, Totowa, NJ
30. Vargha P (1997) A critical discussion of intraclass correlation coefficients. Stat Med 16:821–823
31. Kim DG, Christopherson GT, Dong XN, Fyhrie DP, Yeni YN (2004) The effect of microcomputed tomography scanning and reconstruction voxel size on the accuracy of stereological measurements in human cancellous bone. Bone 35:1375–1382
32. Chappard D, Retailleau-Gaborit N, Legrand E, Basle MF, Audran M (2005) Comparison insight bone measurements by histomorphometry and muCT. J Bone Miner Res 20:1177–1184